LEGIBILITY NOTICE

A major purpose of the Technical Information Center is to provide the broadest dissemination possible of information contained in DOE's Research and Development Reports to business, industry, the academic community, and federal, state and local governments.

Although a small portion of this report is not reproducible, it is being made available to expedite the availability of information on the research discussed herein.

1

Conf-8907136

LA-UR--89-3784

DE90 003212

TITLE THE LARGE COEFFICIENT PROBLEM; CAN WE MAKE SENSE OUT OF QCD PERTURBATION THEORY?

AUTHOR(S): GEOFFREY B. WEST

SUBMITTED TO PROCEEDINGS OF RADIATIVE CORRECTIONS WORKSHOP HELD AT THE UNIVERSITY OF SUSSEX, BRIGHTON, ENGLAND, JULY 1989

Los Alamos National Laboratory
Los Alamos, New Mexico 87545

THE LARGE COEFFICIENT PROBLEM; CAN WE MAKE SENSE OUT OF QCD PERTURBATION THEORY? [1]

Geoffrey B. West

Theoretical Division, T-8
Los Alamos National Laboratory
MS B285
Los Alamos, NM 87545

I INTRODUCTION

There is the possibility of an impending crisis looming on the horizon for QCD. The problem is that in many processes, large coefficients arise in the perturbation series expansion leading to serious uncertainties concerning its predictive power. Until recently most of the examples of such a phenomenon occurred in the calculation of decay rates. These were, by and large, either ignored or dismissed using possible scheme-dependence arguments as a way out. However, more recently a calculation of the 3-loop contribution to the total e^+e^- annihilation cross-section was performed which gave an enormous coefficient of the order of 50 times that of the 2-loop term[(1)]. If correct, this would imply that the 3-loop contribution actually exceeds that of the 2-loop! Thus, **from a conservative viewpoint, the validity of the perturbation series expansion as an estimate for the total e^+e^- cross-section is called into question.** Such a cautionary attitude should even be extended to the lowest order parton-model result, $\sum Q_i^2$; (Q_i being the charge of the ith quark species). Since this process has played a key rôle in the development and understanding of QCD and since, in many ways, it is one of the cleanest methods for extracting α_s (the conventional QCD fine structure constant) the problem can no longer be avoided. Furthermore, there is no reason to doubt (and, in fact, good reasons to believe) that this problem should occur in all physical processes. Coming to grips with it is, of course, not only important for testing QCD but also for extracting fundamental quantities such as α_s. Clearly one needs to understand the nature and origin of such large coefficients before one can confidently continue to use perturbative estimates. Such problems can be expected to occur universally so, using different methods to determine α_s, for example, will not circumvent the difficulty.

The purpose of this talk is to focus on these problems. I shall first review the experimental situation with some examples illustrating the problem. I shall then discuss various general

---

[1] Talk given at the Radiative Corrections Workshop held at the University of Sussex, Brighton, England, July 1989

components and properties of perturbation theory (such as renormalization and causality) before attempting to give a possible resolution of the problem.

Everyone is, of course, familiar with perturbation theory as a calculational tool; Feynman diagrams and their accompanying rules of computation are the stock-in-trade for all particle physicists, theorists and experimentalists alike. Indeed, the phenomenal success of quantum electrodynamics is surely one of the crowning achievements of modern physics. Its predictions are in remarkable agreement with experiment, in some cases to one part in 10^{12}. Calculationally, this success is rooted in perturbation theory which has become the cornerstone for understanding the consequences of any relativistic quantum field theory. In spite of this, perturbation theory has serious limitations beyond those associated, for example, with such questions as bound states, or spontaneous symmetry breaking. Already in 1952 Dyson[2] had pointed out in an elegant paper that the perturbation series could not be convergent. The argument is deceptively simple and very physical: imagine changing α to $-\alpha$ so that like charges now attract and opposite ones repel. Clearly the ground state of this new theory is quite different from that of ordinary QED since virtual pairs created in the vacuum now repel one another. Thus, perturbing around the original "trivial" vacuum of QED will clearly be insufficient to describe this new situation; the structure of the new theory and, in particular, of its vacuum cannot, therefore, be obtained by simply setting $\alpha = -\alpha$ in the Feynman perturbation series. Consequently perturbation theory must be non-analytic in α at $\alpha = 0$ signifying that the series has zero radius of convergence. An amusing historical note to this is that Dyson came to this conclusion[3] after having first claimed that the series was in fact convergent and that QED was therefore a closed book!

If the nature of the divergence of the series were such that it were asymptotic then the situation is, at least in principle, controllable. For, in such a case, as will be reviewed below, a good estimate for the sum of the series is obtained by keeping "only" the first π/α(≈ 400 in QED) terms. In practice, this means that, since α is so small, perturbation theory will, in fact, give an accurate estimate; only at absurdly high order do serious deviations begin to develop. Thus the fact that the series is actually divergent would be of no practical importance. Clearly, then, the nature of the divergence (i.e. whether, for example, it is asymptotic or not) is a potentially deep and important question. It presumably bears upon the question of the self-consistency of QED and whether it needs to be imbedded in a larger, possibly asymptotically free theory.

The second limitation is a practical one and was best expressed by Feynman[4] himself in 1959; he was concerned about developing an approximate algorithm for estimating higher order terms in the perturbation series without having to laboriously (and, to some extent, mindlessly) calculate Feynman diagrams. Of course, his concern is somewhat less of a problem these days given the advent of fast computers and sophisticated software, nevertheless, his remarks are worth repeating. To quote:

"It seems that very little physical intuition has yet been developed in this subject.

> In nearly every case we are reduced to computing exactly the coefficient of some specific term. We have no way to get a general idea of the result to be expected. To make my view clearer, consider, for example, the anomalous electron moment, $[\frac{1}{2}(g-2) = \alpha/2\pi - 0.328\alpha^2/\pi^2]$. We have no physical picture by which we can easily see that the correction is roughly $\alpha/2\pi$, in fact, we do not even know why the sign is positive (other than by computing it). In another field we would not be content with the calculation of the second-order term to three significant figures without enough understanding to get a rational estimate of the order of magnitude of the third. We have been computing terms like a blind man exploring a new room, but soon we must develop some concept of this room as a whole, and to have some general idea of what is contained in it. As a specific challenge, is there any method of computing the anomalous moment of the electron which, on first rough approximation, gives a fair approximation of the α term and a crude one to α^2; and when improved, increases the accuracy of the α^2 term, yielding a rough estimate to α^3 and beyond?"

Although we shall not be able to meet Feynman's challenge directly, nevertheless the techniques discussed here do constitute the beginning of an answer.

Returning to the question at hand, namely QCD rather than QED, we should note that the difficulties there are exacerbated for at least two independent reasons: (i) since $\alpha_s \gg \alpha$, the problem of the divergence of the series is much more serious and (ii) there are explicit non-perturbative phenomena (instantons and the like) associated with new local minima of the action. The question of the interplay between other minima of the action beyond the trivial one and ordinary perturbation theory is a subtle one which we shall discuss below. Regardless, it is clear that the QCD situation is a serious one in that large coefficients can (and do) occur early in the expansion. In some cases, as reviewed immediately below, they occur ridiculously early. Ultimately a methodology based on understanding the nature of the series must be devised for handling them.

II REVIEW OF SOME EXAMPLES

Before discussing some explicit QCD examples, let us examine QED and use it to briefly discuss the question of scheme dependence. One of the best known QED series is that for (g-2) of the electron which was quoted above by Feynman; it reads

$$\frac{1}{2}g_e = 1 + 0.5(\alpha/\pi) - 0.328(\alpha/\pi)^2 + 1.183(\alpha/\pi)^3 + \cdots \tag{1}$$

which certainly looks like a well-behaved series. Recall that in deriving this equation a certain renormalization scheme has been implicitly used; for example, α could be defined through threshold Thompson scattering from (on-shell) electrons. It is via such a definition that we deduce from experiment that $\alpha \approx (137.03\cdots)^{-1}$. This is a natural definition especially for low-energy quantities such as g_e. One could, in principle, use other schemes associated with high energy QED phenomena where the corresponding α would be smaller. Generally speaking, different schemes are related by some polynomial relationship[5]: $\alpha' = \alpha + a_1\alpha^2 + a_2\alpha^3 + \cdots$ (the a_i being constants). Although, the final result for a physical quantity such as g_e does not depend on which scheme is chosen it is clearly not very sensible to choose one associated with a high energy process when dealing with low energy phenomena. In any case, in QED, since electrons are observable, there are "natural" schemes such as via Thompson scattering, which are the most appropriate ones for the definition of α. This is in contradistinction to QCD where there are no analogous "natural" schemes associated with experiments where quarks or gluons are real observables. Because of asymptotic freedom, however, these can be approximated in a variety of high energy experiments such as the total e^+e^- total cross-section measurement. In any case, as already stated, the final result represented, for example, by the sum of the perturbation series must be scheme invariant. On the other hand, to a finite order in perturbation theory, the result will, in general, be scheme-dependent and this is a major source of ambiguity (and confusion). There have been many attempts to define a "best" scheme appropriate to a particular process, however, none are totally satisfactory and all necessarily leave a residue of uncertainty. I shall not, in this talk, be much concerned with such problems especially since for any physical process, **scheme invariant quantities can be defined**[5,7] even in finite orders of perturbation theory. However, to illustrate the problem suppose we use a scheme where[6] $\alpha'/\pi = \alpha/\pi(1 - 10\alpha/\pi)$, then the series (1) reads

$$\frac{1}{2}g_e = 1 + 0.5(\alpha'/\pi) + 4.67(\alpha'/\pi)^2 + 94.61(\alpha'/\pi)^3 + \cdots \qquad (2)$$

which now looks like a badly behaved series! This means that when dealing with large coefficients some care must be taken to express quantities in a scheme-invariant fashion. In this example, incidentally, $(\alpha')^{-1} \approx 140.29$ corresponding to an "inappropriate" α' defined at an energy scale significantly greater than the low energy scale of g_e.

Before moving onto QCD, it is worth mentioning one other example from QED and that is the decay of orthopositronium into three photons. The width for this process is given by[6]

$$\Gamma \approx \frac{2\alpha^6}{9\pi} m_e(\pi^2 - 9)[1 - 10.35(\frac{\alpha}{\pi}) + \cdots] \qquad (3)$$

There are several points worth noting about this; first is the appearance of a large coefficient in the leading order correction; this receives its dominant contribution from the graph shown in fig.1 and is large in any reasonable scheme. Note, however, that, because the tree-graph contribution is $\sim \alpha^6$, this correction is rather sensitive to the scheme. Eq. (3) is one of the few (and possibly only) cases where the theoretical predictions of QED are in serious disagreement with experiment. In fact, there is a 5-standard deviation discrepancy which could be explained if the coefficient of the $(\alpha/\pi)^2$ term were of order 300! Scheme dependence has been evoked to explain this, but it could be a situation where true large coefficients are occurring. Notice, incidentally, the occurrence of the curious coefficient $(\pi^2 - 9)$; conceivably the fact that this almost vanishes (presumably accidentally) contributes to the sensitivity of this process.

Perhaps the first example of a large coefficient occurring in perturbative QCD is in the decay of the 0^{-+} heavy quark state η_B into two gluons[8]. Typical graphs are shown in fig. 2. The calculation yields

$$\frac{\Gamma(\eta_B \rightarrow 2g)}{\Gamma(\eta_B \rightarrow 2\gamma)} = \frac{2}{9Q_B^2}\left(\frac{\alpha_s}{\alpha}\right)\left[1 + 22.4\left(\frac{\alpha_s}{\pi}\right) + \cdots\right] \tag{4}$$

Bound states effects have been completely ignored and the initial valence quarks and final state gluons treated as if free. As in the orthopositronium case, the large coefficient (22.4) is scheme-dependent since the tree-graph contribution is $O(\alpha_s^2)$.

Ruling out a light higgs has been a frustratingly difficult enterprise mostly because of its very weak coupling[9]. The best limits come from rare K and B decays, the higgs being detected via its decay into $\mu^+\mu^-$ pairs. This, therefore, requires accurate knowledge of the branching ratio into this mode which can lead to serious uncertainties. An elegant way of avoiding this problem is to use the decay $\Upsilon \rightarrow h + \gamma$ and to look for a single hard photon in which case, the decay mode of the higgs is irrelevant[10]. At tree level, the decay proceeds via the diagram shown in fig. 3. Ignoring bound state effects and treating the system as if it were positronium, one obtains

$$\frac{\Gamma(\Upsilon \rightarrow h\gamma)}{\Gamma(\Upsilon \rightarrow \mu^+\mu^-)} \approx \frac{G_F m_b^2}{\sqrt{2}\,\pi\alpha}\left(1 - \frac{m_h^2}{m_\Upsilon^2}\right) \tag{5}$$

In this ratio the crude bound state effects, which would be reflected by the wave-function at the origin in a non-relativistic loosely bound system, cancel. In such an approximation $m_\Upsilon \approx 2\,m_b$. In addition to inherently non-perturbative bound state corrections to this formula

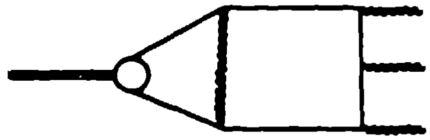

Fig. 1

Dominant contribution to the decay of orthopositronium into three photons

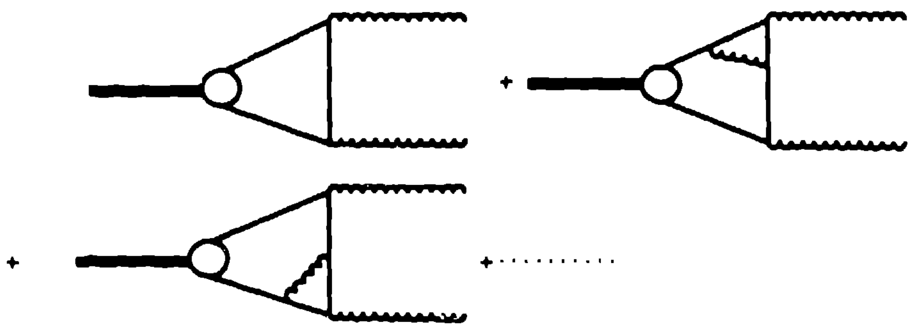

Fig. 2

Typical graphs contributing to the decay of paraquarkonium

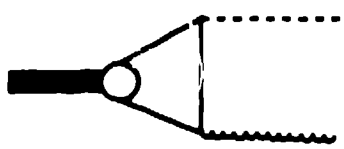

Fig. 3

Leading graph in the decay $\Upsilon \dashrightarrow h + \gamma$

there are the usual QCD perturbative radiative corrections to the sub-process $b\bar{b} \rightarrow h\gamma$. A calculation of these leads to[11]

$$\Gamma(Y \rightarrow h\gamma) \approx \Gamma_o(Y \rightarrow h\gamma)\left[1 - a\left(\frac{\alpha_s}{\pi}\right)\right] \tag{6}$$

where, for $m_h \ll m_Y$, $a \approx 13$ again showing the appearance of a large coefficient. Taken at face value this formula would rule out [12] a higgs in the range 200 MeV$\lesssim m_h \lesssim$6GeV. However the correction here reduces the tree graph result by ~ 80 % making the calculation suspect. When uncertainties about possible bound state corrections are folded in, one must certainly take this conclusion *cum grano salis*, especially since a conservative attitude is mandated when dealing with the existence of the higgs! Notice incidentally that unlike the previous formulae the expansion in eq. (6) begins with $(\alpha_s)^\circ$ so that the result should be scheme-invariant.

This process has already been alluded to in the introduction. The total cross-section (R) for e^+e^- annihilation into hadrons relative to that of $\mu^+\mu^-$ pairs is directly related to the absorptive part of the polarization tensor defined by

$$\Pi_{\mu\nu}(q) = i\int d^4x e^{iq\cdot x}\langle 0|T[j_\mu(x)j_\nu(0)]|0\rangle \tag{7}$$

$$= [q^2 g_{\mu\nu} - q_\mu q_\nu]\Pi(q^2) \tag{8}$$

where $j_\mu(x)$ is the electromagnetic current operator: R = 12πIm Π. If μ is an arbitrary renormalization scale parameter then the perturbative expansion of R reads

$$R\left[q^2/\mu^2, \alpha_s(\mu)\right] = \left(\sum Q_i^2\right)\sum_{n=0}^{\infty}\tilde{a}_n(q^2/\mu^2)(\alpha_s/\pi)^n. \tag{9}$$

In the $\overline{MS}$ scheme with 5 flavors, the coefficients have the following [1] values when $q^2 = |\mu^2|$:

$$\tilde{a}_o = \tilde{a}_1 = 1\ ;\ \tilde{a}_2 = 1.41 \text{ and } \tilde{a}_3 = 64.9. \tag{10}$$

The last of these is truly remarkable. To get some idea of its implications, note that if it is neglected then a comparison with data at $\sqrt{q^2} = 34\,GeV$ leads to $\alpha_s = 0.169$ corresponding to $\Lambda_{QCD} \approx 610\,MeV$. On the other hand, it if is included then $\alpha_s = 0.150$ corresponding to $\Lambda_{QCD} \approx 314\,MeV$, a reduction of 50%. To make this even more dramatic, Fleischer et al [13] have made a Padé approximant fit to the series. Though this should be taken *cum grano* *salis* it does illustrate just how serious things could become: they find $\tilde{a}_4 \approx 10^4$ leading to $\alpha_s \approx 0.114$ and $\Lambda_{QCD} \approx 60.3\,MeV$ and $\tilde{a}_5 \approx 1.5 \times 10^6$ leading to $\alpha_s \approx 0.086$ and $\Lambda_{QCD} \approx 6.3\,MeV$. Clearly, then, one needs to understand the nature of the series and its implicit non-perturbative character before one can begin to confidently extract quantities like α_s and $\sum Q_i^2$ from the data. In a sense one can compare this situation with that which existed in weak interactions prior to unification. At that time the Fermi theory gave an adequate (and reasonably accurate) description of low energy weak interaction phenomenology in spite of the fact that it was non-renormalizable so that eventually it would break down. Similarly, here, one might argue that, because of asymptotic freedom, perturbation theory should give an adequate estimate at truly infinite energies. However, at finite energies appropriate to present-day experiment, there are potentially large corrections due to the divergent nature of the series, even though the theory is renormalizable. Just as one had to wait for the development of a renormalizable theory of the weak interactions in order to control and consistently define the infinities in each term of the perturbative expansion, so, from a conservative point of view, one must await a similar procedure for dealing with the complete sum of the series before being confident that our predictions are consistently meaningful.

III GENERAL PRINCIPLES, DEFINITIONS AND TECHNIQUES

Perturbation theory can be thought of as being generated by an expansion of the path integral around local minima of the action [S_m, say, where $\delta S/\delta A_\mu|_{s=s_m} = 0$]. Symbolically, then, Π can be thought of as having the following representation:

$$\Pi(q^2, g^2) \approx \sum_{m,n=0}^{\infty} A_{mn}(q^2) e^{-Sm/g^2} (g^2)^{n-\nu_m} \qquad (11)$$

where g is the usual gauge coupling ($\alpha_s \equiv g^2/4\pi$) and the dependence on μ has been suppressed. The series with $m = 0$ defines ordinary perturbation theory as represented by the usual sum of Feynman graphs. The multi-instanton sectors and so forth, which are topologically separated from ordinary perturbation theory, are represented by the $m \neq 0$ series. As generated directly from the path integral, the expansion (11) is very sick: each coefficient $A_{mn}(q^2)$ is divergent as is each series summed over n. The first of these problems is conventionally dealt with via some renormalization scheme (and, if necessary, by some infrared cut-off) whereas the second disease is typically ignored. The challenge is to find a consistent scheme to control the divergence implicit in the sums. To attack this problem, I shall need to review some general properties of series expansions with emphasis on asymptotic expansion. However, before doing so I want to remind the reader of the constraints on Π rendered by renormalizability and causality. The point is that the latter dictates general analytic properties as a function of q^2 whereas the former tells us that q^2 and g^2 are not, in fact, independent variables. Thus, the general dependence on, and analytic structure, in g^2 is, in some sense, known.

Because j_μ is a composite operator there is an additional divergence above the usual multiplicative ones needed to renormalize QCD that must be cancelled to render the theory finite. This is the single $q\bar{q}$ intermediate state which gives a logarithmically divergent contribution to Π even in free field theory. On the other hand j_μ, being a conserved current, has no anomalous dimension. The renormalization group equation resulting from the invariance of Π to changes in the scale μ therefore reads (neglecting quark masses)[14]

$$\left[\mu\frac{\partial}{\partial\mu} + \beta(g)\frac{\partial}{\partial g}\right]\Pi\left[\frac{q^2}{\mu^2}, g^2(\mu)\right] = I[g^2(\mu)] \tag{12}$$

The presence of I reflects the composite nature of j_μ. Now, the general solution to this equation can be expressed as follows[14]:

$$\Pi\left(\frac{q^2}{\mu^2}, g^2\right) = F\left[\frac{q^2}{\mu^2}e^{2K(g)}\right] + \phi(g^2) \tag{13}$$

where

$$K(g^2) \equiv \int^g \frac{dg'}{\beta(g')} \tag{14}$$

$$\text{and} \quad \phi(g^2) \equiv \int^g dg'\frac{I(g')}{\beta(g')}, \tag{15}$$

F being an arbitrary function.

In perturbation theory

$$\beta(g) \approx -g^3(b_1 + b_2 g^2 + \cdots) \tag{16}$$

leading to

$$K(g) \approx \frac{1}{2b_1 g^2}\left[1 + \frac{b_2}{b_1}g^2 ln\left(\frac{1}{g^2} + \frac{b_2}{b_1}\right) + 0(g^4)\right] \tag{17}$$

Now, both R and $D \equiv (q^2 \partial/\partial q^2)\Pi$ satisfy the homogeneous equation and so depend only on the single variable $z \equiv (q^2/\mu^2)e^{2K(g)}$. Thus, for such quantities, $q^2 \rightarrow \infty$ is equivalent to $g^2 \rightarrow 0^+$ This, of course, is just the asymptotic freedom connection, namely that the asymptotic q^2 behavior of R is deriveable from its small g^2 behavior and so, can presumably be systematically calculated via perturbation theory.

Conversely, it is clear that the small q^2 behavior is equivalent to $g^2 \rightarrow 0^-$. Thus, if perturbation theory [i.e. eq. (9)] were convergent, so that D or R were analytic in g^2 at $g^2 = 0$, then one would have proven a remarkable theorem (15), namely, that their infra-red and ultra-violet behaviors had to be identical! Put slightly differently one could state this as saying that the difference between the IR and UV behaviors reflects the lack of analyticity at $g^2 = 0$. Indeed, if one knew the precise nature of the singularity at $g^2 = 0$ then one would know the IR behavior of the theory! Thus the problem of the large coefficients in QCD is presumably linked to the problem of its IR behavior.

It is well-known that causality implies that Π be an analytic function of q^2 for complex q^2 except for possible singularities along the positive real axis. This is normally expressed via a dispersion relation. Now, asymptotic freedom dictates that for $q^2 \rightarrow \infty$, $\Pi \sim \ln q^2$ thereby requiring (at least) one subtraction(14). This subtraction in the dispersion relation is, in fact, intimately related to the extra subtraction needed to renormalize Π and, consequently, to the inhomogeneity in the renormalization group equation. Using this it is easy to write a dispersion representation for D:

$$D\left(\frac{q^2}{\mu^2}, g^2\right) = \frac{q^2}{12\,\pi^2}\int_0^\infty \frac{dq'^2}{(q'^2 - q^2)^2}\, R\left(\frac{q'^2}{\mu^2}, g^2\right) \tag{18}$$

$$= \frac{q^2}{\mu^2}e^{2K(g)}\int_0^\infty \frac{dz}{\pi}\frac{f(z)}{\left[z - \frac{q^2}{\mu^2}e^{2K(g)}\right]^2} \tag{19}$$

In writing the second line, the RG constraint that both D and R be functions of z only [i.e. $R = 12\,\pi^2 f(z)$] have been explicitly incorporated.

Notice that no assumption about a mass gap has been made here. For this amplitude, the appearance of a mass gap (beginning at $4\,m_\pi^2$) is related to the introduction of quark masses. For the glueball channel, however, where a similar representation holds, it is generally expected that there exists a mass gap even in the massless quark limit(16). In such a case the corresponding D will have a Taylor series expansion in q^2:

$$D(q^2/\mu^2, g^2) = \sum_{n=0}^{\infty} d_n(g^2)(q^2)^n \tag{20}$$

This expansion is the complement to the perturbation series where the expansion is in powers of g^2 with q^2-dependent coefficients. However, in contrast to that expansion which is asymptotic at best, this expansion necessarily has a finite radius of convergence (given by the square of the glueball mass, or, in the case we have been considering, $4m_\pi^2$ if quark masses are introduced).

Requiring that D be a function of z only determines the full g^2-dependence of the d_n:

$$D(q^2/\mu^2, g^2) = \sum_{n=0}^{\infty} \tilde{d}_n \left[q^2/\mu^2 e^{2K(g)}\right]^n \tag{21}$$

$$\simeq \sum_{n=0}^{\infty} \tilde{d}_n \, e^{n/b_1 g^2} \left(\frac{1}{g^2} - \frac{b_2}{b_1}\right)^{nb_2/b_1^2} \left(\frac{q^2}{\mu^2}\right)^n \tag{22}$$

where $\tilde{d}_n$ are numbers independent of q^2, μ^2 or g and, in the second line, we have used eq. (17). The question that must now be faced is how can this (exact) expression, which contains explicit non-analytic pieces in g^2 - indeed, essential singularities - ever be cast into the form of a perturbative expansion in g^2, as in eq. (9), which naïvely treats D as if it were analytic in g^2 at $g^2 = 0$? In order to begin to answer this question we need to digress a little into some properties of series expansions.

IV DIGRESSION ON SERIES EXPANSIONS

Consider the following series expansion

$$f(x) = \sum_{n=0}^{\infty} \frac{(-1)^n}{n!} a_n x^n \tag{23}$$

I first want to invert this to obtain a formula for the a_n in terms of $f(x)$. The trick is to make use of properties of the gamma-function[17] $\Gamma(s)$, namely, that it has a string of simple poles at $s = -n$ with residue $(-1)^n/n!$ Thus, (23) can be expressed as a contour integral

$$f(x) = \int_c \frac{ds}{2\pi i} \Gamma(s) a(s) x^{-s} \tag{24}$$

where $a(s)$ is the analytic continuation of a_n such that $a(-n) \equiv a_n$ and C is the contour shown in the fig. 4. Now, if $a(s)$ has no singularities in the left-hand plane and the integrand

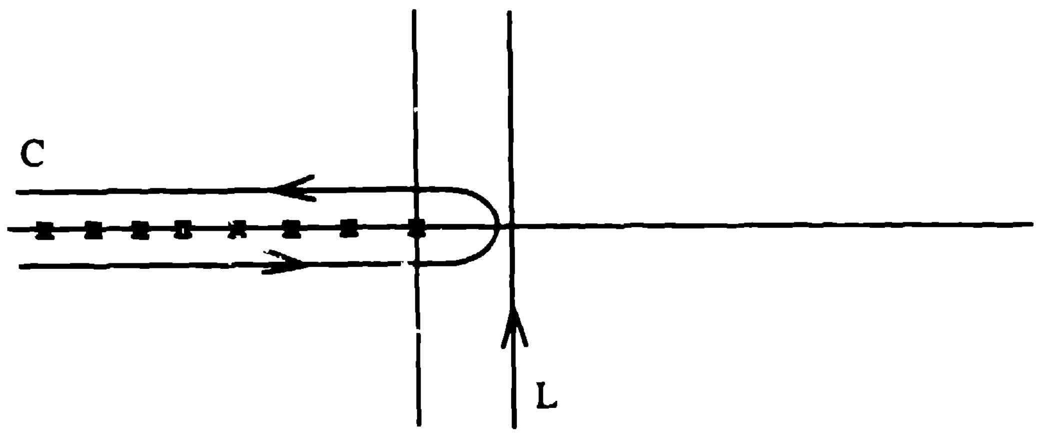

Fig. 4

Complex s-plane showing singularities of Γ (s) and the contour C used in eq. (24). C can be deformed to L and a Mellin transforms profound to obtain eq. (25)

is sufficiently convergent, C can be replaced by a line L parallel to the imaginary axis. The resulting representation will be recognized as an inverse Mellin transform from which one can read off the desired result:

$$a(s) = \frac{1}{\Gamma(s)} \int_0^\infty dx\, x^{s-1} f(x) \tag{25}$$

An alternative form for $a(s)$ that is sometimes useful can be obtained by analytically continuing the integrand into the complex x−plane:

$$a(s) = \Gamma(1-s) \int_c \frac{dx}{2\pi i} (-x)^{s-1} f(x) \tag{26}$$

The contour C wraps around the cut defined along the positive real axis necessary to define x^{s-1} as a single-valued function. It can be opened up to pick up the singularities of $f(x)$. For example, if these only occur on the negative real axis then (26) reads

$$a(s) = \Gamma(1-s) \int_0^\infty \frac{dx}{\pi} x^{s-1} Im\, f(-x) \tag{27}$$

Suppose the series (23) were divergent; define a new series $g(x)$ from it by introducing coefficients b_n such that

$$g(x) = \sum_{n=0}^{\infty} \frac{(-1)^n}{n!} a_n b_n x^n \tag{28}$$

This new series can always be made convergent by choosing the b_n to fall sufficiently fast for large n. Using the Mellin transform technique described in (A) above together with its convolution theorem leads to the following formula:

$$f(x) = \int_0^\infty \frac{du}{u} g(u) B(\frac{x}{u}) \tag{29}$$

$$\text{where } B(x) \equiv \int_L \frac{ds}{2\pi i} \frac{x^{-s}}{b(s)} \tag{30}$$

i.e. $1/b(s)$ is the Mellin transform of $B(x)$. As before $b(s)$ is the analytic continuation of b_n such that $b(-n) \equiv b_n$.

The idea of summability is then the following: choose the b_n to make the series for $g(x)$ convergent; insert its sum into (29) which, if it exists, gives a well-defined representation for $f(x)$. The best-known version of this is due to Borel[17]: choose $b_n = 1/n!$ then $B(s) = 1/\Gamma(1-s)$ leading to $B(x) = (1/x)e^{-1/x}$. This defines the Borel sum of $f(x)$:

$$f(x) = \frac{1}{x}\int_o^{\infty} du\, g(u)e^{-u/x} \tag{31}$$

Other choices for the b_n are, of course, possible, however the Borel technique is the one that has received most attention. As an example of the Borel method consider the series generated by $a_n = (n!)^2$; this is clearly divergent. However, $g(x) = \sum(-x)^n$ can be summed to give $(1+x)^{-1}$ and so

$$f(x) = \int_o^{\infty} \frac{dv\, e^{-v}}{1+xv} \tag{32}$$

This is supposed to be the true unique representation of $f(x)$. From this point of view the original divergent series simply arose from our "illegal" expansion of the integral as a power series in x.

The question arises as to when this technique does, in fact, give a unique and consistent representation of the function. There are, naturally, many important theorems and treatises dealing with such questions; however, this is not the place to review them. Roughly speaking, the method works when all the integrals converge uniformly. Of particular importance is the absence of singularities on the positive real u-axis. For example, if $a_n = (-1)^n(n!)^2$ then $g(u) = (1-x)^{-1}$ and the series is no longer Borel summable. Typically this means that undetermined essential singularities such as $e^{-a/x}$ cannot be excluded from $f(x)$.

Suppose that $f(x)$ has a power series expansion in some wedge of analyticity in the complex plane $\theta < \pi/2$. Consider

$$\left| f(x) - \sum_{n=0}^{N} \frac{(-1)^n}{n!} a_n x^n \right| \equiv R_N(x) \tag{33}$$

For the sorts of series that we are interested in $R_N(x) \lesssim C_N x^{N+1}$ so that $R_N(x)/x^N \to \infty$ when $N \to \infty$ for x fixed indicating zero radius of convergence. On the other hand $R_N(x)/x^N \to 0$ for N fixed and $x \to 0$. This is Poincaré's definition of an asymptotic series. As shall be demonstrated below, a general feature of quantum field theories is that $C_N \sim cb^{N+1}\Gamma(N+a)$ where a, b, and c are constants. It is easy to confirm that $R_N(x)$ minimizes when

$$N = N_o \approx 1/bx - a \tag{34}$$

and that

$$R_{N_o}(x) \approx c(2\pi bx)^{\frac{1}{2}} e^{-1/bx} \tag{35}$$

This is a remarkable result which demonstrates the character of asymptotic series for it shows that $R_N(x) \to 0$ for x sufficiently small. **Thus when $N = N_o$, the series exponentially approaches the correct value of the function $f(x)$ for sufficiently small x even though it diverges! If further terms are added to the partial sum one is driven further from the correct result.** Thus, if one believes that in QED the appropriate expansion parameter is α/π and that the series is asymptotic in Poincaré's sense then it is not until π/α terms that one need be concerned! Furthermore one can approach within $e^{-\pi/\alpha}$ of the exact result!

V LARGE n-BEHAVIOR

As explained in the Introduction the main thrust of this paper is to gain some possible insight into the occurrence of large coefficients in the perturbation expansion. The question of summability discussed above, though intimately related to this problem, will be discussed elsewhere. The rest of the paper is therefore devoted to the question of the large n behaviour of the coefficients $a_n(q^2/\mu^2)$. We shall first review how this can be attacked using the "bare" path integral representation. Unfortunately this leaves several questions unanswered, especially for gauge theories[18]. We therefore turn to the representation (19) which incorporates q^2-analyticity and renormalizability and therefore, implicitly, the complete g^2-dependence. Furthermore, in contrast to the path integral it is a representation for the truly physical amplitude.

With a generalization to the path integral in mind, consider functions $f(x)$ which have the following representation

$$f(x) = \frac{1}{\sqrt{x}} \int_{-\infty}^{\infty} du e^{-A(u)/x} \tag{36}$$

For example, if $A(u) = u^2 + u^4$ then

$$f(x) = \int_{-\infty}^{\infty} du e^{-(u^2 + xu^4)} \tag{37}$$

which is the "zero-dimensional" limit of a Euclidean ϕ^4 field theory with coupling strength x. Notice that if this is naïvely expanded in powers of x, one obtains

$$f(x) \approx \sum_{n=0}^{\infty}(-1)^n \frac{\Gamma(2n+\frac{1}{2})}{\Gamma(n+1)} x^n \tag{38}$$

which is clearly divergent. Indeed it can be Borel-summed to reconstruct the original representation (37). Inserting (36) into the coefficient generating formula, eq. (25), gives

$$c(s) = \frac{\Gamma(\frac{1}{2}-s)}{\Gamma(s)} \int_{-\infty}^{\infty} du\, e^{(s-\frac{1}{2})\ln A(u)} \tag{39}$$

Thus $\ln A(u)$ generates the coefficients. Let us suppose, again with our eye on field theory, that $A(u) \sim u^2$ when $u \to 0$, then by continuing u into the complex plane, it is possible to re-express (39) in the form

$$a(s) = \frac{\Gamma(\frac{1}{2}-s)\Gamma(1-s)}{\cos \pi s} e^{2\pi i s} \int_C \frac{du}{2\pi i} e^{(s-\frac{1}{2})\ln A(u)} \tag{40}$$

where the contour C wraps around the cut on the positive real axis necessary to define $\ln u$. This expression is ripe for exploitation by the method of steepest descents. Saddle points occur when $[\ln A(u)]' = 0$ i.e. $A'(u) = 0$. Notice that, although $A'(u) = 0$ when $u = 0$, this is not so for $[\ln A(u)]'$; thus, even though $u = 0$ is a saddle point of the original representation of $f(x)$ and is the point about which perturbation theory is developed, it is not a saddle point of the coefficient generating function, eq. (40). Typically $[\ln A(u)]'' > 0$ at the saddle point (u_0) and we find that for $s = -n \to -\infty$

$$a(-n) = a_n \approx \frac{[\Gamma(1+n)]^2}{(n+1)} \frac{[-A(u_0)]^{-n}}{[2\pi A''(u_0)]^{\frac{1}{2}}} \tag{41}$$

provided $A(u_0) < 0$ which is generally valid for polynomial $A(u)$. It is straightforward to check that for the example (37), this formula agrees with eq. (38). **Eq. (41) shows that the effective expansion parameter is actually not x, but rather $x/A(u_0)$.** Furthermore, note that if $A(u_0) > 0$ then, naïvely, a factor $(-1)^n$ is induced in (41) which would imply that the series is no longer Borel summable. We shall return to this situation below.

The extension of the above analysis is straightforwardly generalizeable to a path integral representation. The vacuum-vacuum amplitude for a scalar field theory with action functional $A[\phi]$ is given by

$$W(g) = \frac{1}{\sqrt{g}} \int \mathcal{D}\phi e^{-A[\phi]/g} \tag{42}$$

where g is the coupling constant. This can be expanded in the usual Feynman graph perturbation series as a power expansion in g. The coefficients can be determined, as above, using eq. (25):

$$a(s) = \frac{\Gamma(1-s)\Gamma(\frac{1}{2}-s)}{2\pi i \cos \pi s} e^{2\pi i s} \int \mathcal{D}\phi e^{(s-\frac{1}{2}) \ln A[\phi]} \tag{43}$$

Thus $ln\ A[\phi]$ acts as the effective action for determining the a_n. As before the trivial local minimum of A at $\phi = 0$, though being the starting point for perturbation theory, does not contribute to $a(s)$. The saddle points (at $\phi = \phi_o$, say), satisfy the classical equations of motion and have an action given by

$$A[\phi_o] = -\int d^4x \phi_0^4 = -8\pi^2/3 < 0 \tag{44}$$

The functional integral can be evaluated at the saddle point and an answer analogous to (41) derived. Care must be taken in properly accounting for zero-modes etc. with the result that

$$a_n \approx n[\Gamma(n+1)]^2 \{-A_o[\phi_o]\}^{-n-\frac{1}{2}} \tag{45}$$

Thus the expansion parameter is not g but rather $(3g/8\pi^2)$.

In the literature this formula was originally derived using eq. (27) for $a(s)$[18]. This requires an immediate analytic continuation in g. Now for $Re\ g < 0$, eq. (42) diverges indicating that singularities occur only in the left-hand plane. To determine the nature of these singularities, the path integral itself needs to be analytically continued in ϕ. One finds a cut beginning at $g = 0$ extending along the negative $Re\ g$ axis. An evaluation of the discontinuity across this cut gives a result in agreement with eq. (45). Deriving the result this way makes a connection with Dyson's original argument since an imaginary part only develops if there are other vacua that are not stable.

In attempting to extend this technique to non-abelian gauge theories such as QCD serious problems arise. First, there is the classic problem of maintaining gauge invariance for physical quantities. Secondly, these theories lead to non-trivial saddle-points with **positive** action. As already emphasized this precludes a straightforward application of summability. On the other hand these additional minima of the action (typically referred to as instantons and the like) have a topological characterisitc associated with them. In that sense they give rise to a more general expansion beyond ordinary perturbation theory in terms of topological sectors, as represented in eq. (11). The problem is then to determine how much, if any, of the instanton-like contributions feed back to what is usually thought of as ordinary perturbabion theory. Because of problems such as these it has been difficult to apply these techniques directly to the path integral representation.

The representation, eq. (19), incorporates both causality and renormalizability and, as such, explicitly contains information that is not directly encoded in the path integral. In this sense it is potentially more useful for our purposes since it contains an essential feature of perturbation theory, absent in the path integral, namely renormalizability. In this section I shall therefore attempt to exploit (19) to determine the large n behavior of the coefficients.

Let us first express the perturbative expansion of D in the form

$$D\left(\frac{q^2}{\mu^2}, g^2\right) \approx \sum_{n=0}^{\infty} \frac{(-1)^n}{n!} A_n \left(\frac{q^2}{\mu^2}\right) (g^2)^n \tag{46}$$

The coefficients that we are actually interested in are those occurring in the expansion of R as in eq. (9). These can be derived from the A_n via the formula:

$$\tilde{a}_n = \left(-4\pi^2\right)^{n+1} \frac{3}{\pi n b_1} \left[\frac{\mathrm{Im}\ A_{n+1}}{(n+1)!} + \frac{b_2}{b_1} \frac{\mathrm{Im}\ A_n}{n!} + \cdots\cdots\right] \tag{47}$$

For large n only the first term need be kept since we anticipate that $A_n \sim (n!)^2$. In order to avoid apparent essential singularities at the origin, arising from the renormalization group, it is convenient to transform to the variable $k \equiv 1/g^2$. Using eqs. (26) and (19) the coefficients in (46) can be obtained from

$$A\left(s, \frac{q^2}{\mu^2}\right) = \Gamma(1-s) \frac{q^2}{\mu^2} \int_0^{\infty} \frac{dz}{\pi} f(z) \int_c \frac{dk}{2\pi i} \frac{(-k)^{-(1+s)} e^{2K}}{[z - q^2/\mu^2 e^{2K}]^2} \tag{48}$$

with K given by eq. (17). We are interested in the behavior of this expression when $s \to -\infty$. As before, this can be estimated using a steepest descents technique. The structure of the complex k-plane is evidently quite complicated as can be seen from fig. 5. There are three distinct types of singularity: (i) the familiar cut on the positive real axis necessary to define $(-k)^{-s-1}$; (ii) an infinite sequence of poles, (at $k = k_N$, say) arising from the vanishing of

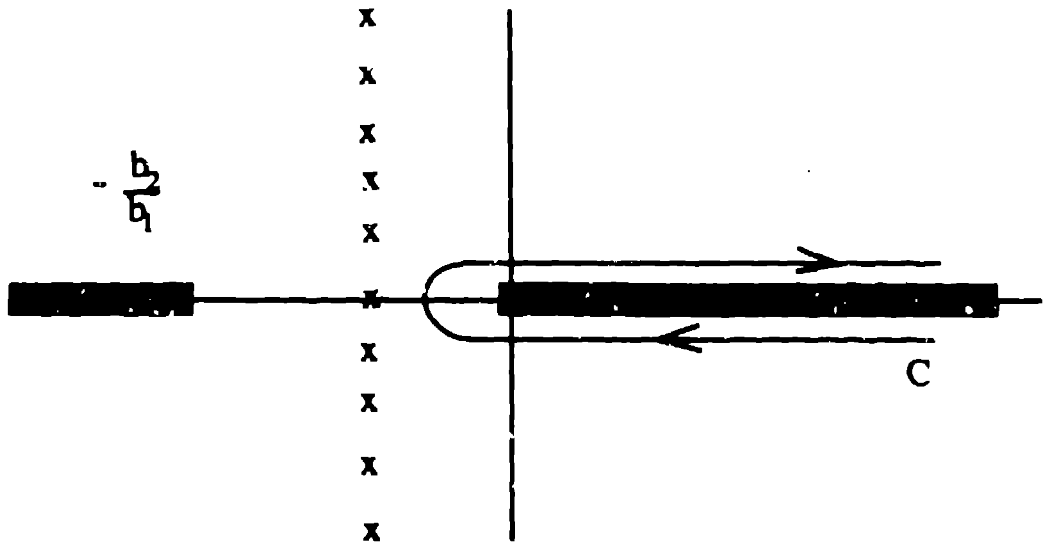

Fig. 5

Complex k-plane showing the contour C used in eq. (48) together with the singularity structure the denominator: $k_N/b_1 + \cdots\cdots \approx \ln(q^2 z/\mu^2) \pm 2\pi N i (N = 0, 1 \cdots)$ and finally (iii) cuts necessary to define potential logarithms in e^{2K}; (e.g. keeping only the first two terms in (16) or (17), there is a cut at $k = -b_2/b_1$ as shown in fig. 5).

As one might guess, this complex structure gives rise to a plethora of saddle points in the k-plane making an accurate estimate of (47) quite subtle. A more detailed discussion of this will be given in a later paper; however, roughly speaking, these saddle points fall into three categories that correspond to the three categories of singularity mentioned above. These occur at (a) $k \approx b_1(s+1) + b_2/b_1 + 0(e^{-s})$ (b) $2K \approx \ln(q^2 z/\mu^2) \pm 2\pi i N + 0(1/s)$; (c) $\beta(g)/g^3 \approx 0$. Let us discuss each of these briefly. The first is the saddle point that we are most interested in, for it dominates the large s behavior. Notice that it corresponds to $g^2 \to 0^-$, as anticipated earlier. The second reflect the poles at $k = k_N$ in (ii) above. They generate the typical $(\ln q^2/\mu^2)^n + \cdots$ dependence of the coefficients familiar from asymptotic freedom. Since we want to compare with ref. 1 where $q^2 = \mu^2$ these saddle points are presumably not of interest here. Finally, there can be saddle points arising from other possible fixed points of $\beta(g)$. Notice that the usual "trivial" fixed point at $g^2 = 0$ (about which perturbation theory is actually generated) is excluded from this. This is analogous to the situation we encountered above when dealing with the path integral formulation where the trivial saddle point at $A = 0$ does not explicitly contribute to the estimate. On the other hand, the "pseudo" fixed point at $k = -b_2/b_1$ generated by keeping only the first two terms, must be included. In QCD the signs of b_2 and b_1 are such that this occurs for Re k < 0 (or Re $g^2 < 0$). Had the sign of b_2/b_1 been different as in ϕ^4 theory (but not in QED!) then this cut migrates to the positive real axis overlapping the $(-k)^{-1-s}$ cut. In such a case it is not obvious that the coefficients can be determined and it is reasonable to speculate that this is related to the well-known claim that ϕ^4 is in fact a trivial theory.

Returning to eqs. (47) and (48) we can estimate the large n-behavior of $\tilde{a}_n$ by keeping only the single saddle- point at $k \approx b_1(s+1)$. In that case we find, assuming $|s| \gg b_2/b_1$, that

$$\tilde{a}_n \sim -\frac{e}{\pi}(4\pi^2 e b_1)^{n-1}\frac{\Gamma(n)}{n^2} \tag{49}$$

Thus

$$\frac{\tilde{a}_{n+1}}{\tilde{a}_n} \approx -4\pi^2 e b_1 n \tag{50}$$

$$\approx -\frac{e}{4}(11 - 2/3N_f)n \tag{51}$$

indicating rapid growth of the coefficients with n. (In these expressions e is the base of natural logarithms, not to be confused with the electronic charge!) Perhaps the most striking aspect of this result is that **the effective expansion parameter is not α/π but rather** $\alpha_{eff} \equiv 4\pi^2 e b_1 \alpha/\pi$. Repeating the analysis of Section IV C for the case here we find that the remainder R_n minimizes when $n \sim 1 + \alpha_{eff}^{-1} + (2 + \alpha\, eff)^{-1} \sim 4.5$ [taking α itself to be

~ 0.15]. This is remarkable for it says that **an accurate result can be obtained by keeping only these first few (4-5) terms**. Indeed, an estimate of the error introduced by this process gives a contribution of only $\sim 10^{-3}$! This is all very encouraging; however, how accurately can we trust these estimates if n is so small? To get some idea, if we put n=2 in eq. (50) then $\bar{a}_3/\bar{a}_2 \sim 11$. This is indeed a relatively large number, although not big enough to account for the result of ref. 1. Furthermore, our asymptotic formula requires that succeeding terms alternate in sign, a characteristic which does not show up in 1. On the other hand the sign of $\bar{a}_3$ agrees with our prediction, so the "problem" resides in $\bar{a}_2$. One certainly would not expect our analysis to be valid for this coefficient so there is no serious contradiction. The problem is, of course, that even though corrections to (49) coming from expanding around the saddle point can be expected to be quite small, there are many other sub-asymptotic saddle points whose contribution may well be comparable to the leading contribution expressed in eq. (49). A more accurate analysis is therefore required to actually establish a firm estimate of $\bar{a}_3$, for example, and to confirm the calculated result. Such an enterprise is currently being undertaken. It is worth noting that eq. (49) gives $\bar{a}_3 \sim 12$, a factor 2π smaller than the calculated number of ref. 1.

It should also be pointed out that these leading estimates are both gauge and scheme-invariant, as one might expect. Ultimately one would like to be able to confidently extract α_s from the data (if it is sufficiently accurate!) which means that we need to know either where to stop the series or how to resum it. Our analysis indicates that stopping at $n \sim 4$ is sufficient. In that case one could simply add the estimate to the already calculated numbers. Similarly one can resum the series beyond these terms using a variant of the Borel technique discussed in the previous Section. In any case it is clear that some consistent procedure or algorithm must eventually be invoked to control the divergence problem and the consequent large coefficients. In this talk I have attempted to show how this problem can be solved in principle and suggested some practical possibilities. A later paper will present details and pursue the solution.

References

[1] S. G. Gorishny, A. L. Kataev and S. A. Larin, Phys. Lett. **212B**, 238 (1988)

[2] F. Dyson, Phys. Rev. **85**, 631 (1952)

[3] F. Dyson, Phys. Rev. **83**, 608 (1951)

[4] R. P. Feynman, Solvay Conference 1959

[5] P. M. Stevenson, Phys. Rev. **D23**, 2916 (1981); C. J. Maxwell, Phys. Rev. **D28**, 2037 (1983); S. J. Brodsky, G. P. Lepage and P. B. Mackenzie, Phys. Rev. **D28**, 228 (1983).

[6] W. Celmaster and D. Sievers, Phys Rev. **D23**, 227 (1981)

[7] A. Dhar, Phys. Lett. **128B**, 407 (1983)

[8] R. Barbieri et al. Nuc. Phys. **B154**, 535 (1979)

[9] See, e.g., S. Raby, G. B. West and C. Hoffman, Phys. Rev. **390**, 828 (1989)

[10] F. Wilczek, Phys. Rev. Lett. **40**, 279 (1978)

[11] M. I. Visotsky, Phys. Lett. **97B**, 159 (1980); P. Nason, ibid. **175B**, 233 (1986)

[12] J. Lee-Franzini, Proc. XXIV Int. Conf. on High Energy Physics (Springer-Verlag, Berlin, 1989) p. 1432

[13] J. Fleischer et al., Univ. of Bielefeld preprint BI-TP 05/89

[14] See, e.g., G. B. West, Nuc. Phys. **B288**, 444 (1987)

[15] G. B. West, Phys. Lett. **145B**, 103 (1984)

[16] G. B. West, Nuc. Phys. B (Proc. Suppl.) **1A**, 57 (1987

[17] E.T. Whittaker and G. N. Watson, "A Course in Modern Analysis". Cambridge Univ. Press, 1950)

[18] For a review see J. Zinn-Justin, Phys. Rep. **70**, 109 (1981).